

Ciência de Dados e Inteligência Artificial

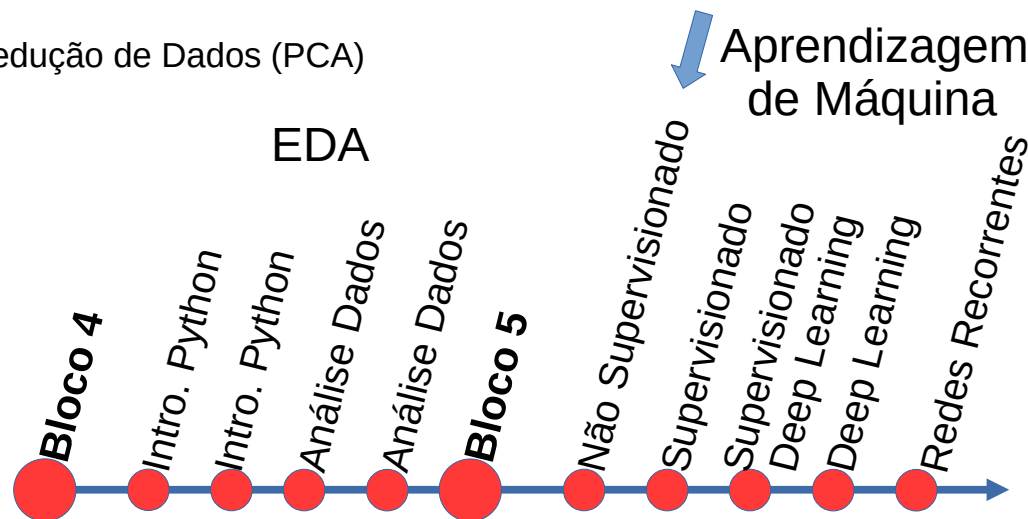
Tópico 03 – Aprendizagem de Máquina Não - Supervisionado

Prof. André G. Hochuli

Tópico 03 – Aprendizado Não Supervisionado

- Programa de Aula

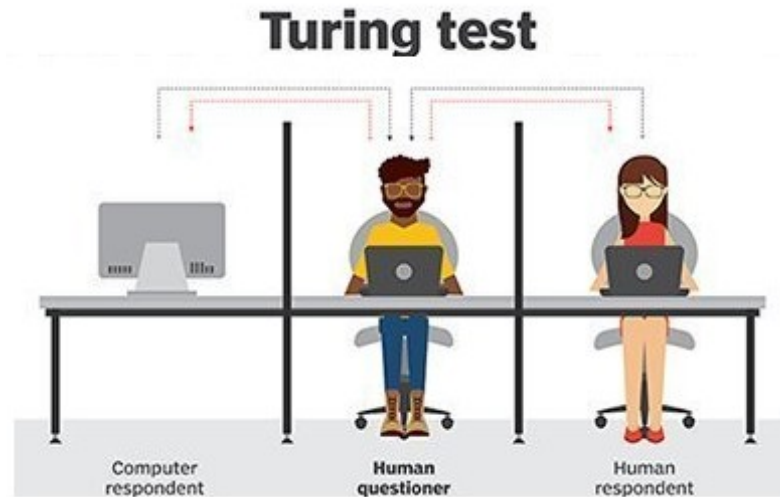
- Inteligência Artificial e Aprendizagem de Máquina
- Tipos de Aprendizado (Sup, N-Sup, Reforço)
- Algoritmos K-Means
- Análise e Métricas
- Normalização e Redução de Dados (PCA)



Inteligência Artificial

“Tópico da Ciência da Computação cujo objetivo é desenvolver **soluções tecnológicas que simulam o raciocínio humano**, ou seja, a inteligência que é característica dos seres humanos.”

“Máquinas podem pensar?” - **Alan Turing** (1950)



IA passa a ser considerada ciência em 1956 (Dartmouth College, USA).

Inteligência Artificial

A inteligência artificial já faz parte do nosso cotidiano há algum tempo.

- Filtros de SPAM
- Sistemas de Recomendação
- Reconhecimento de Padrões
 - Reconhecimento Facial em Imagens
 - Biometria Ocular e Impressão Digital
 - Análise e Monitoramento de Video
 - Inspeção Industrial

Inteligência Artificial

Mais recentemente, dado os modelos profundos, o uso de I.A se tornou mais evidente.

Google creates new self-driving car company



Dave Lee
North America technology reporter

🕒 13 December 2016



GETTY IMAGES

Google has a car that operates without the need for a steering wheel

China embraces artificial intelligence-driven agriculture, emerges as global technology hothouse

Craig Moran | Real Clear Science | July 12, 2018



Image credit: ADC Forum

Inteligência Artificial

Mais recentemente, dado os modelos profundos, o uso de I.A se tornou mais evidente.

AI is acquiring a sense of smell that can detect illnesses in human breath

Computers can take minutes to autonomously analyse a sample that previously took hours to test

Andrea Soltoggio | Wednesday 13 June 2018 12:57 | 0 comments



Artificial intelligence that can teach? It's already happening



ABC Science | By Carl Smith for the Science Show

Updated 15 June 2018 at 10:10 pm

First posted 15 June 2018 at 5:04 pm



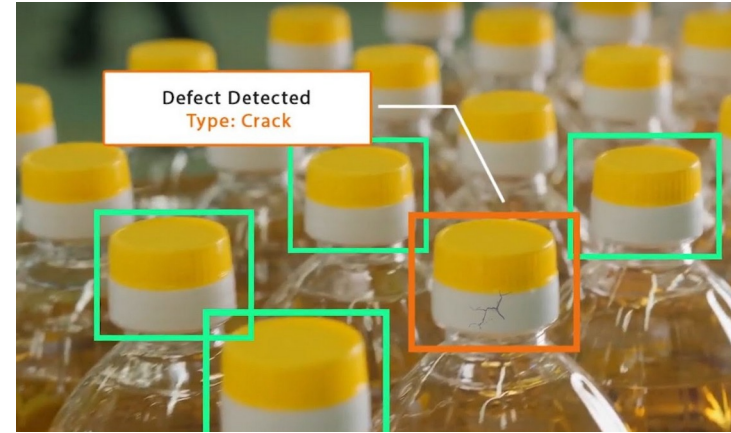
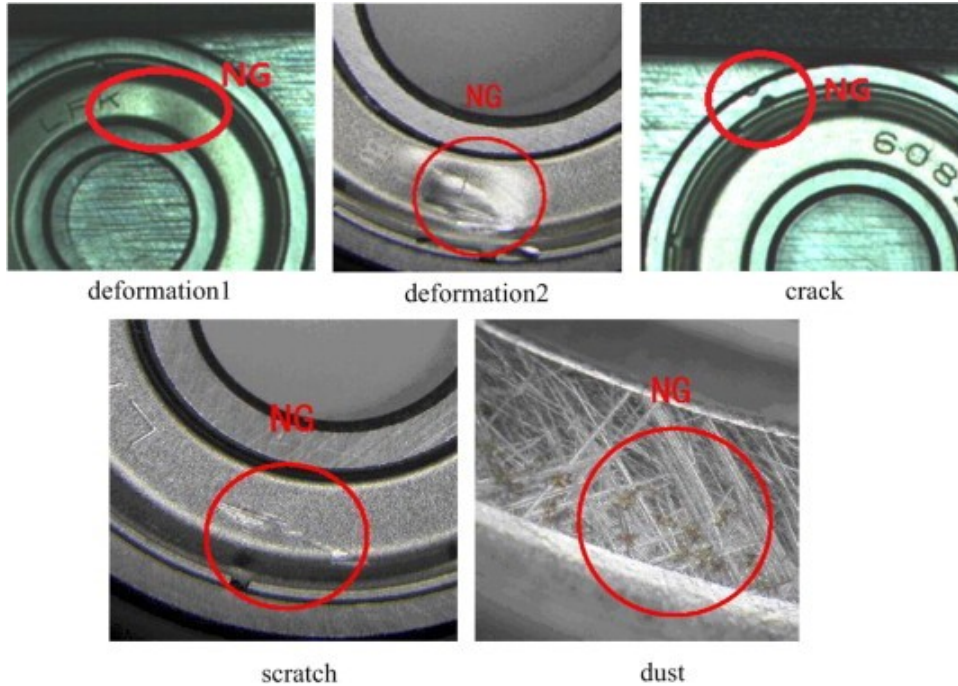
Inteligência Artificial

Mais recentemente, dado os modelos profundos, o uso de I.A se tornou mais evidente.



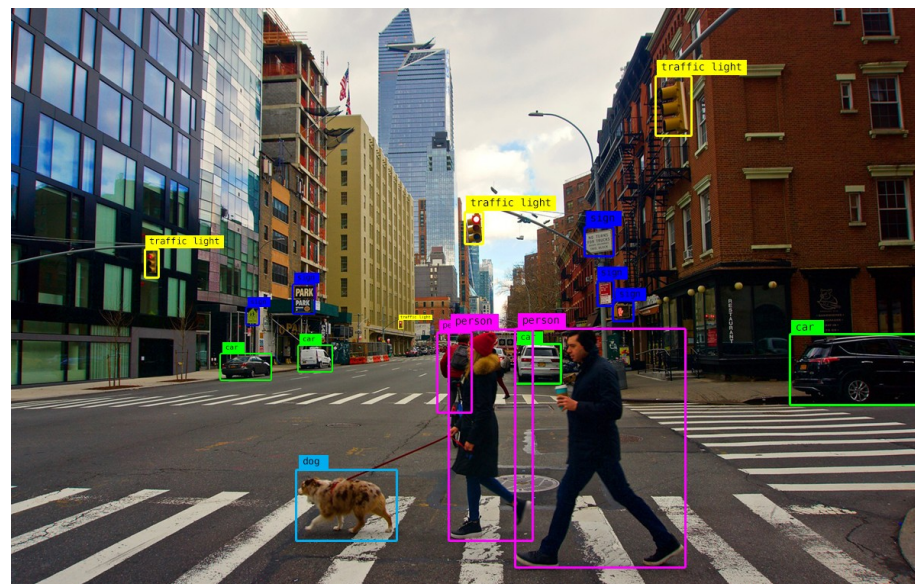
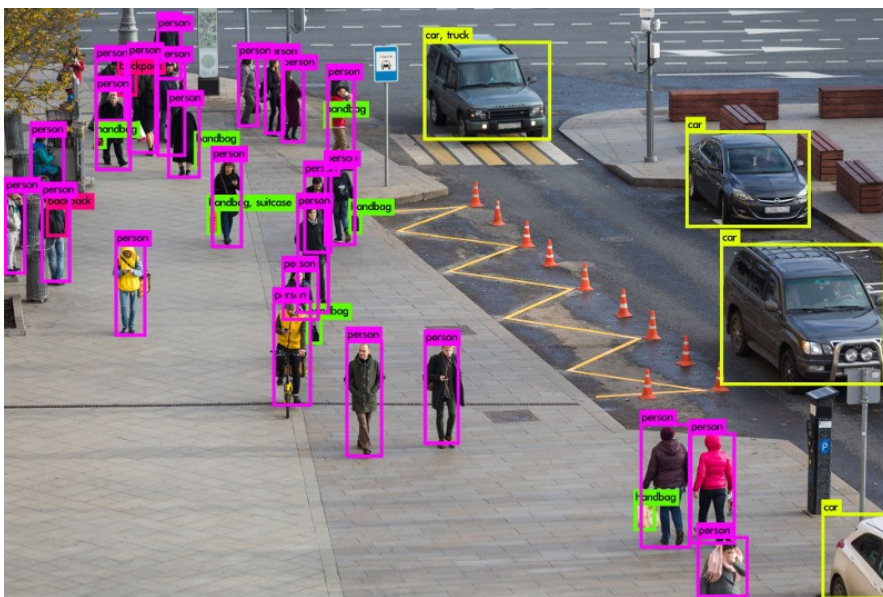
Inteligência Artificial

Mais recentemente, dado os modelos profundos, o uso de I.A se tornou mais evidente.



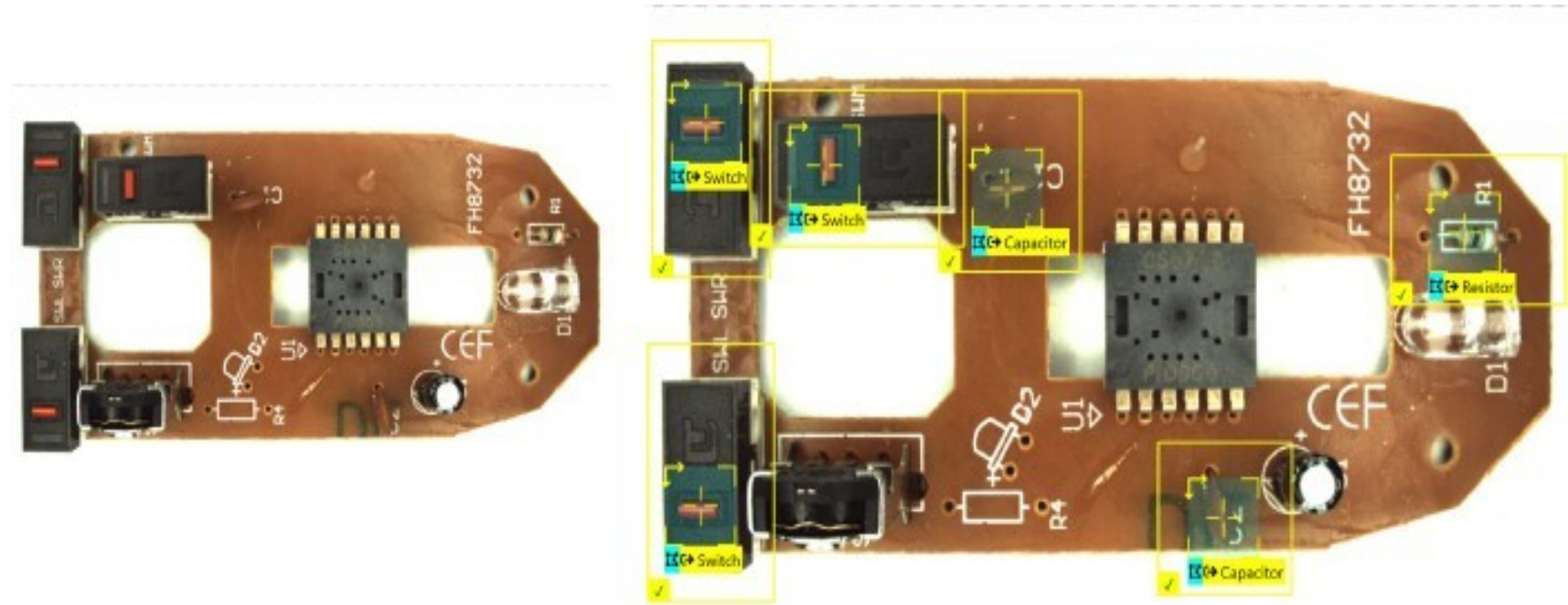
Inteligência Artificial

Mais recentemente, dado os modelos profundos, o uso de I.A se tornou mais evidente.



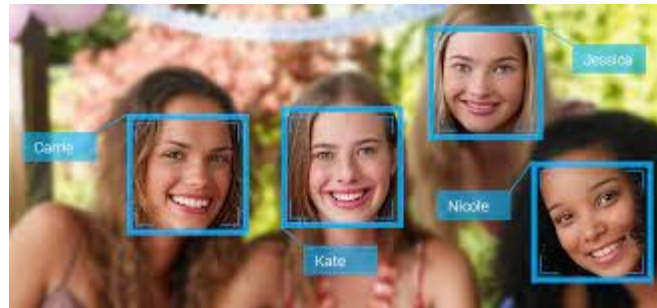
Inteligência Artificial

Mais recentemente, dado os modelos profundos, o uso de I.A se tornou mais evidente.



Inteligência Artificial

Mais recentemente, dado os modelos profundos, o uso de I.A se tornou mais evidente.

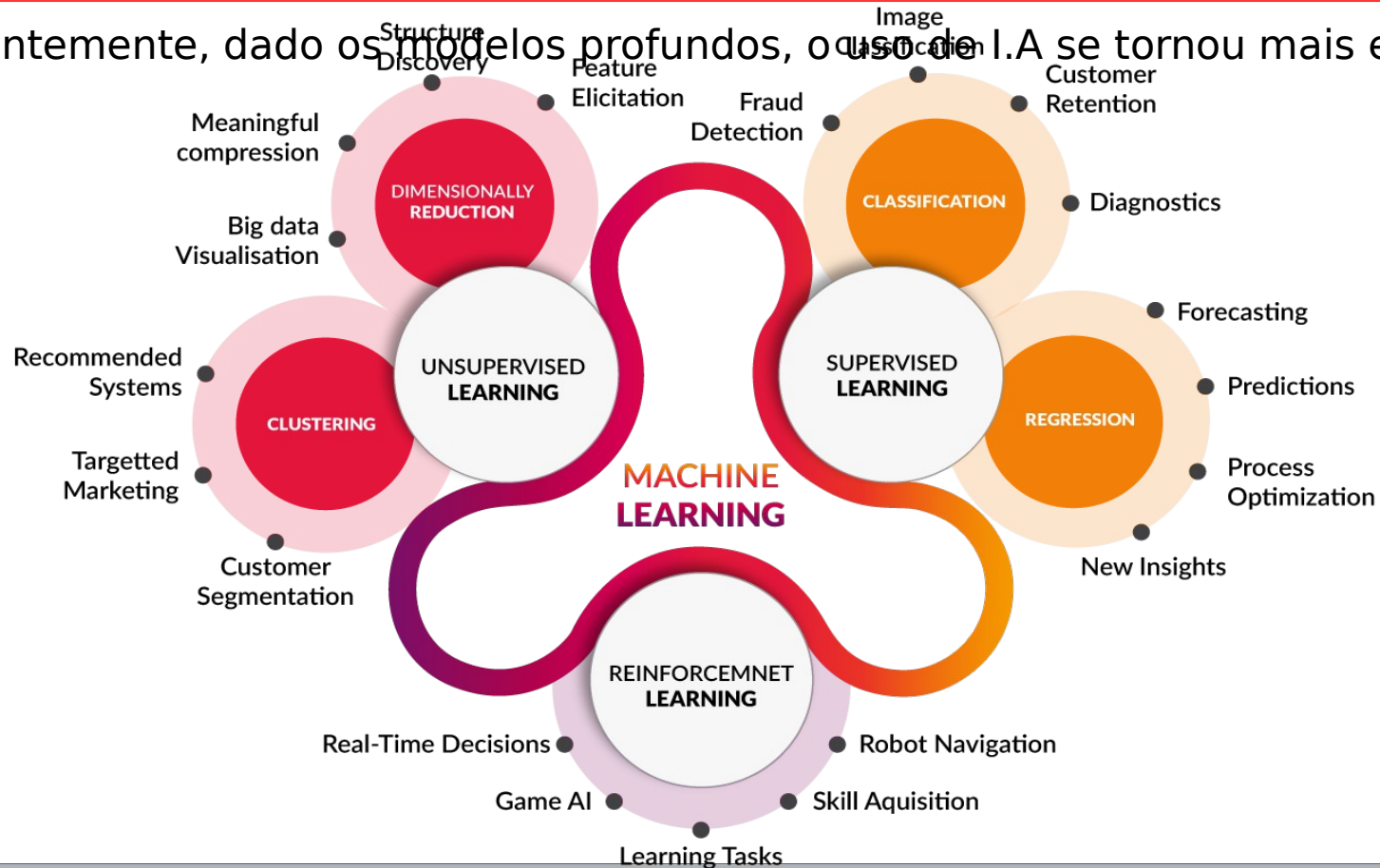


Inteligência Artificial



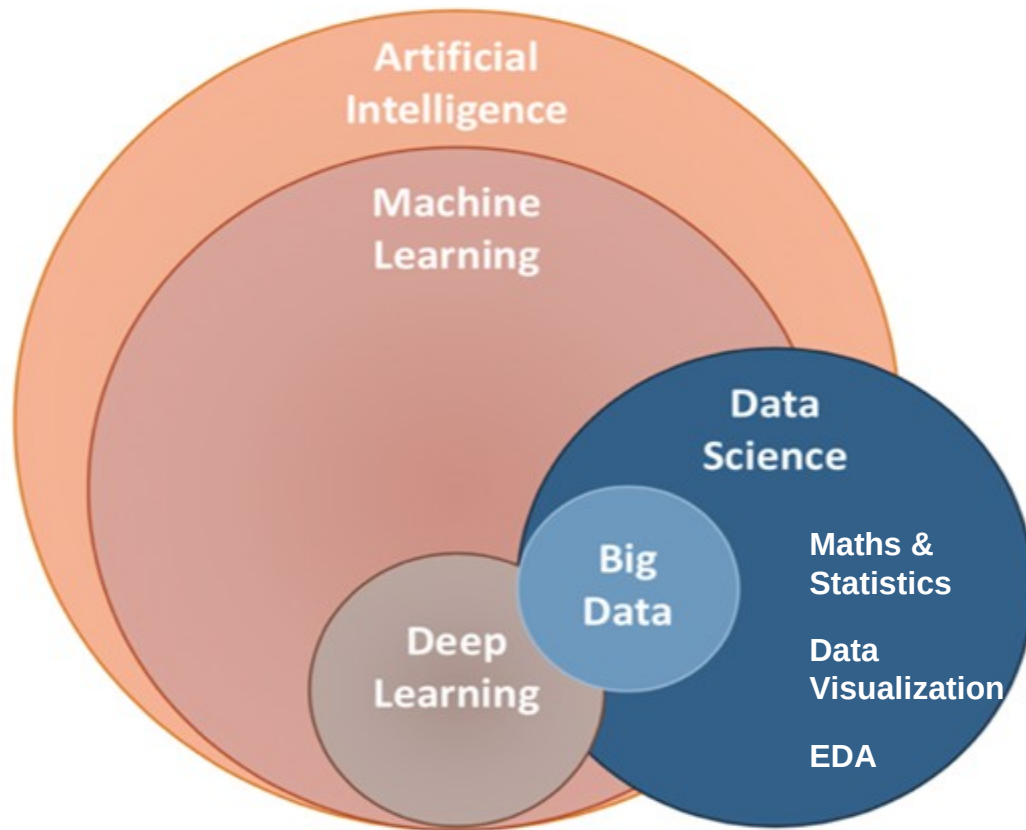
Aprendizado de Máquina

Mais recentemente, dado os modelos profundos, o uso de I.A se tornou mais evidente.



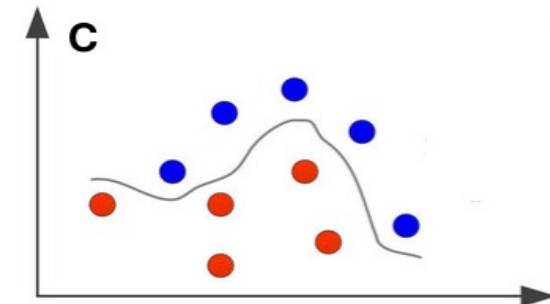
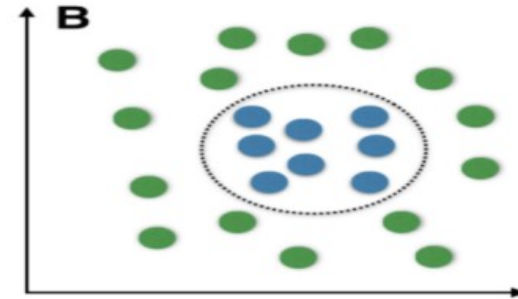
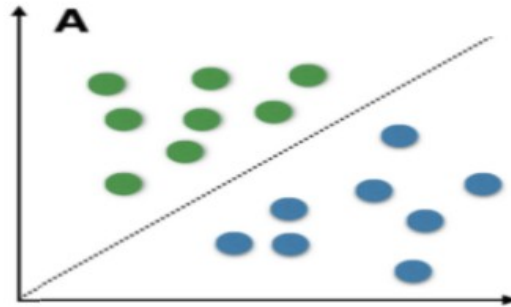
Aprendizado de Máquina

Temas Emergentes



Quando aplicar?

- Representação do problema não é linearmente separável
- A solução não é determinística
- Complexidade Alta (Multi-Dimensões)



Representação

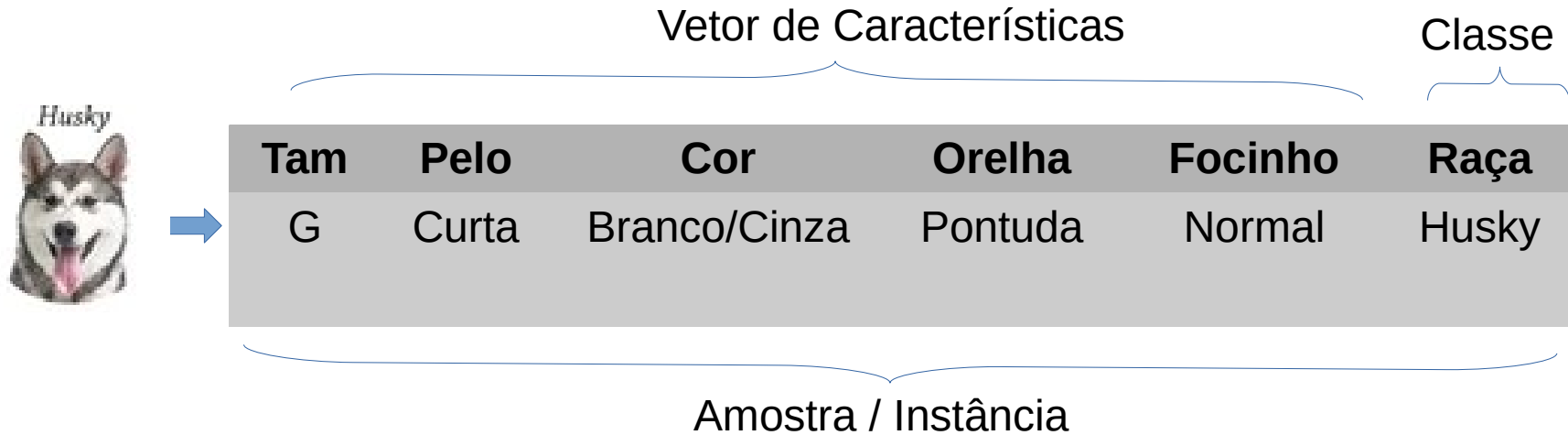
Mundo Real → Dados



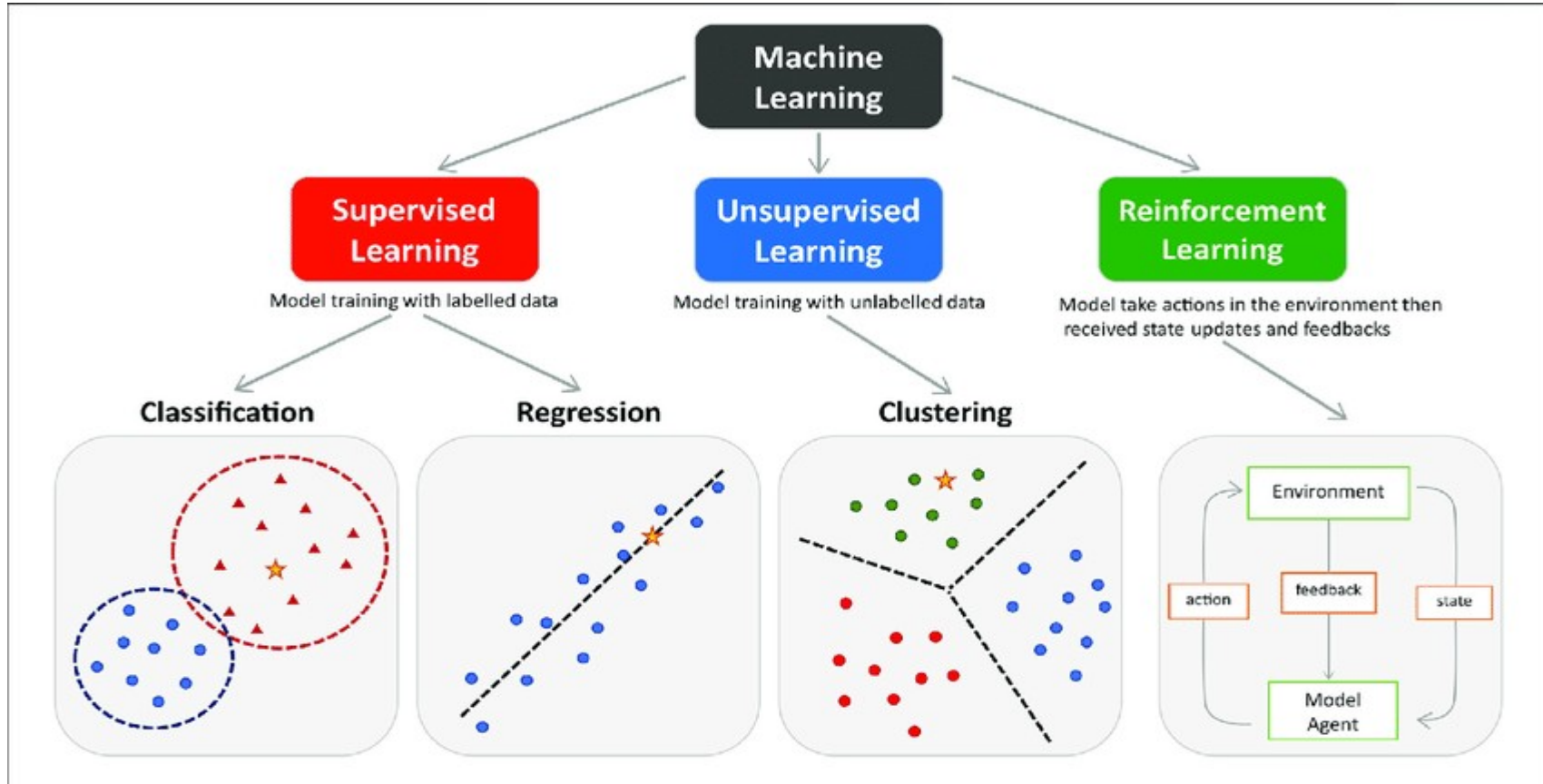
<u>Tam</u>	Pelo	Cor	Orelha	Focinho	Raça
G	Curta	Branco/ Cinza	Pontuda	Normal	Husky
P	Curta	Branco/ Preta	Caída	Achatado	Pug
P	Curta	Caramelo	Pontuda	Normal	Chihuahua
M	Curta	Branco/ Caramelo	Caída	Normal	Beagle
P	Longa	Preta/ Caramelo	Pontuda	Normal	Yorkshire
G	Longa	Caramelo	Pontuda	Normal	Pastor Alemão
G	Curta	Branco/ Caramelo /Preta	Caída	Normal	Labrador

Representação

Terminologias

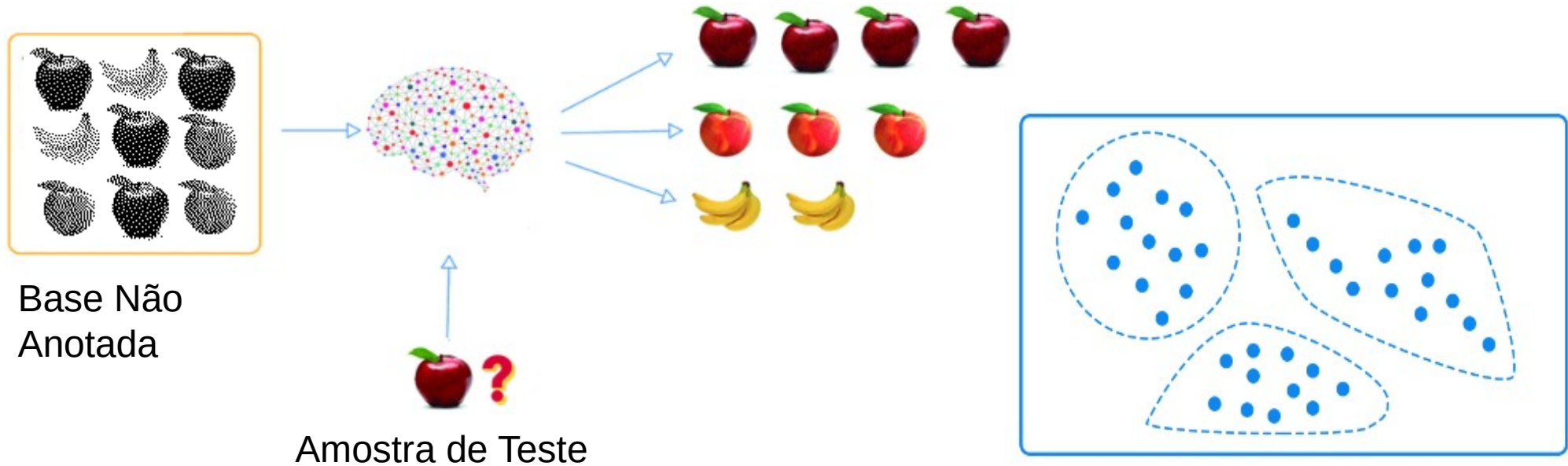


Tipos de Aprendizado



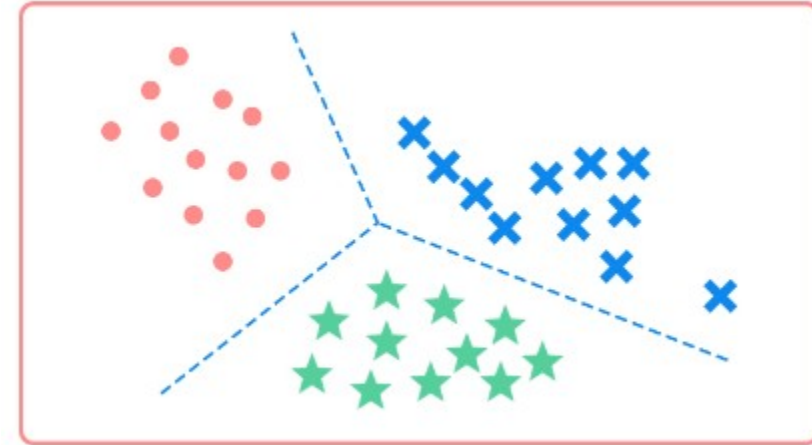
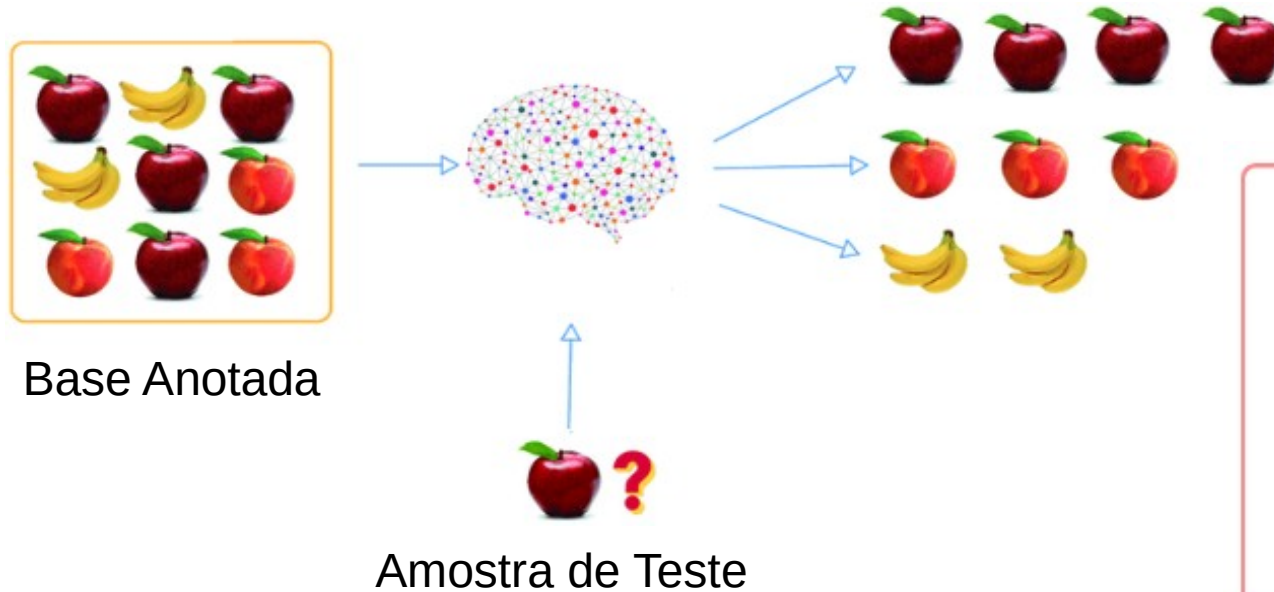
Aprendizado Não - Supervisionado

- Dados Não Anotados (sem categorias)
- Treinamento do Modelo Preditivo
- Agrupamento (Clustering)



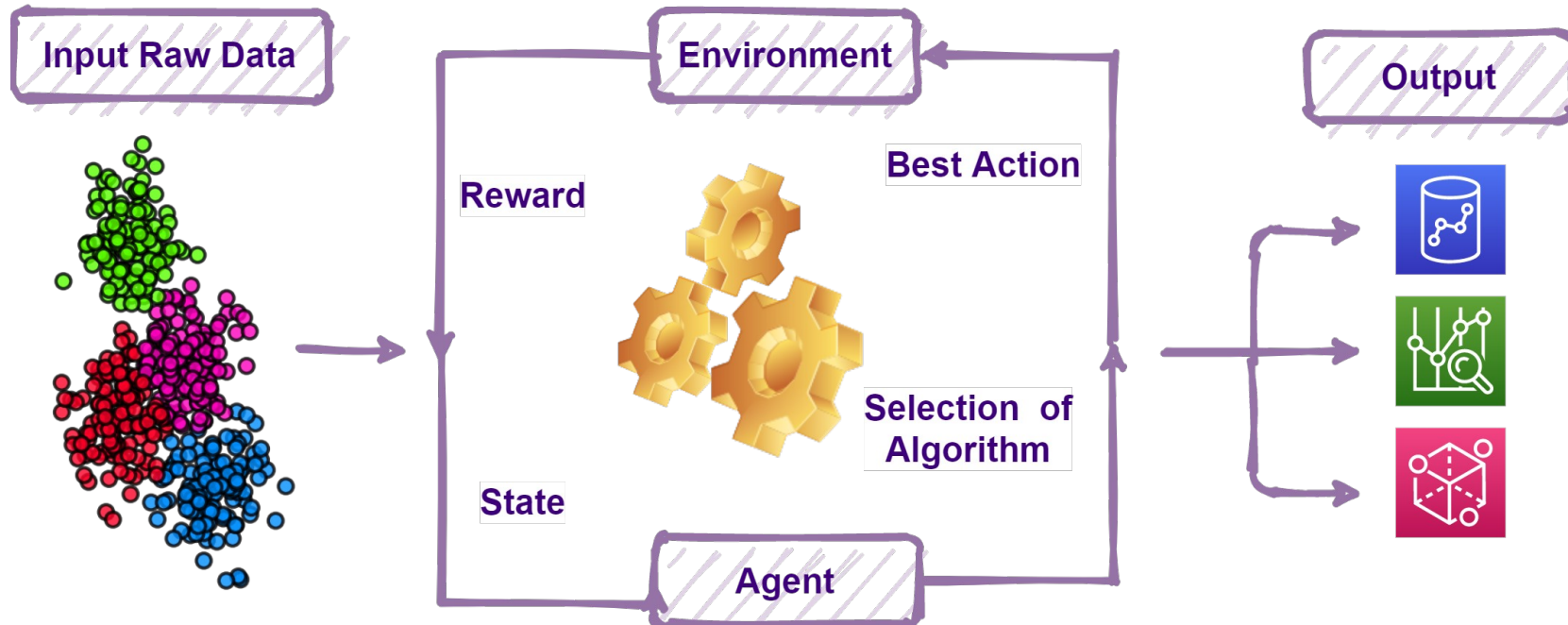
Aprendizado Supervisionado

- Dados Anotados (Categorias)
- Treinamento do Modelo Preditivo
- Classificação / Regressão



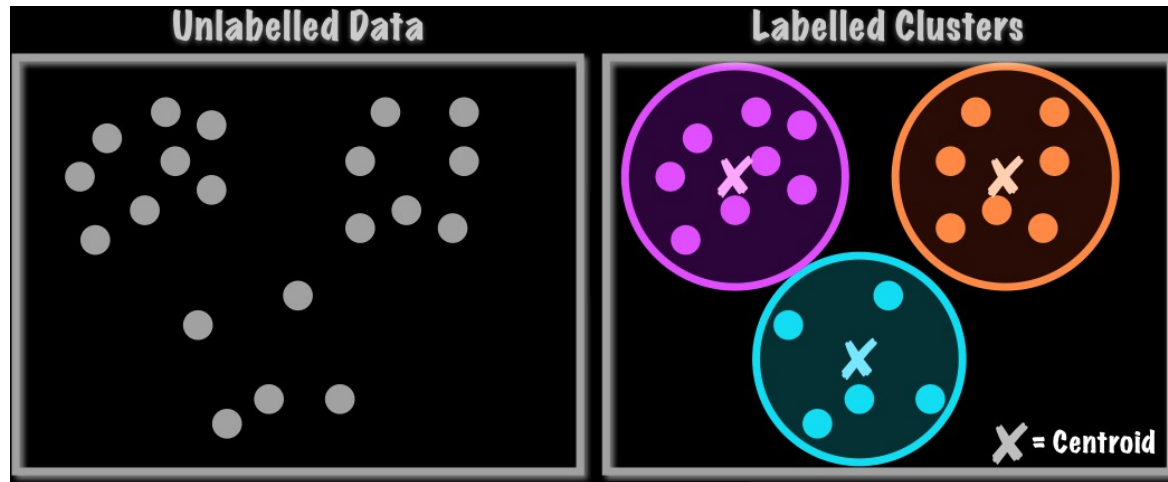
Aprendizado por Reforço

- Dados Anotados / Não anotados
- Interação vs Recompensa



Agrupamento dos Dados (Clustering)

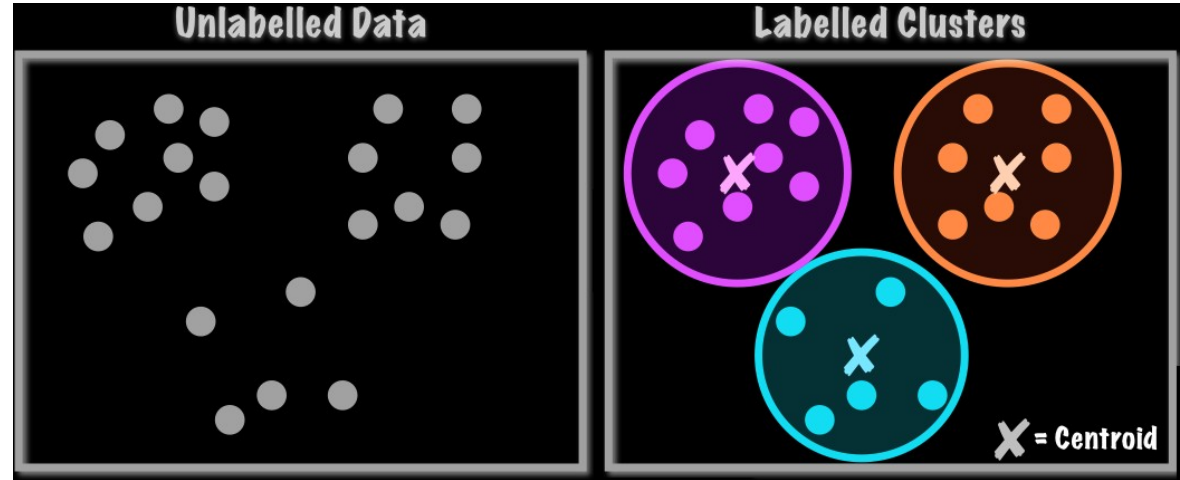
- Encontrar relacionamentos intrínsecos dos atributos
- Métrica de agrupamento determina os grupos (clusters)
- Determinar o número correto de grupos nem sempre é claro



Agrupamento dos Dados (Clustering)

- Aplicações:

- Segmentação de consumidores
- Classificação de Anomalias
- Defeitos, Doenças, Fraudes
- Classificação de espécies
- Organização de Documentos
- Mineração de dados
- Redução de Dimensionalidade

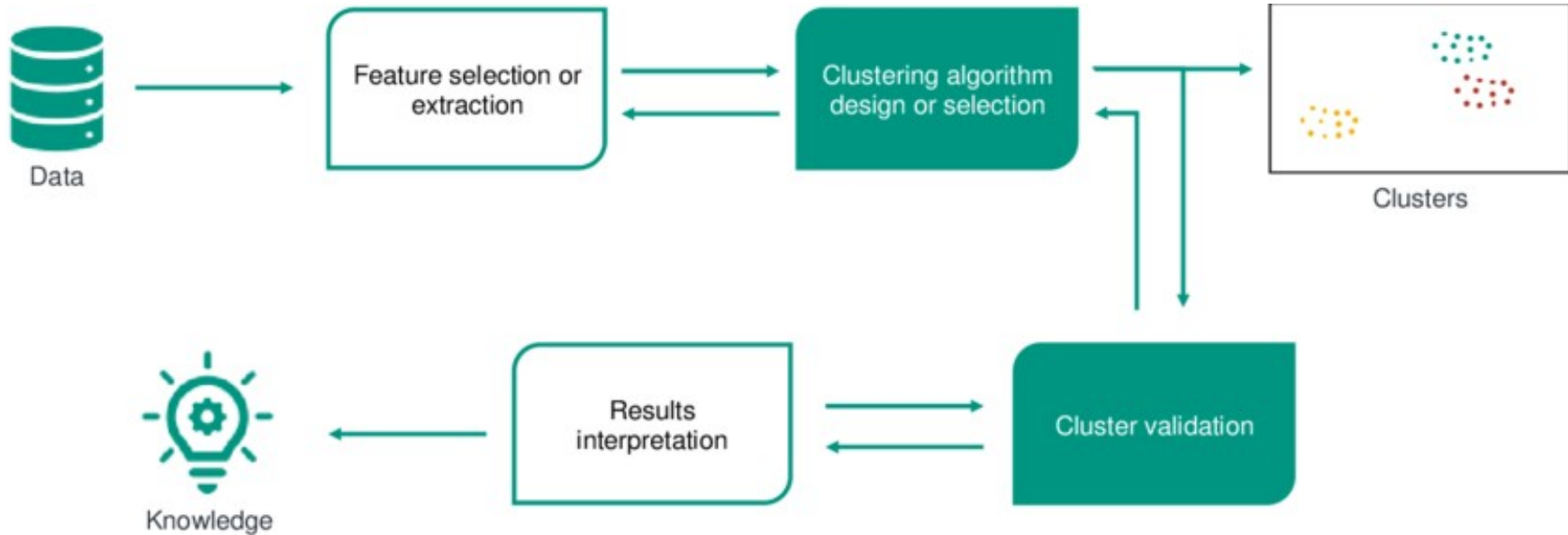


- Casos reais:

- Determinação de Recall baseado em histórico de reparos
- Características de produtos não vendidos (i.e 220v em cidades 110v)

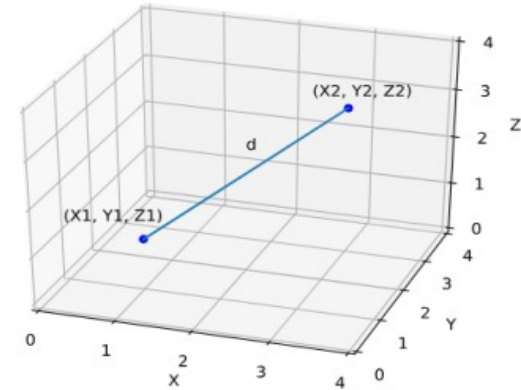
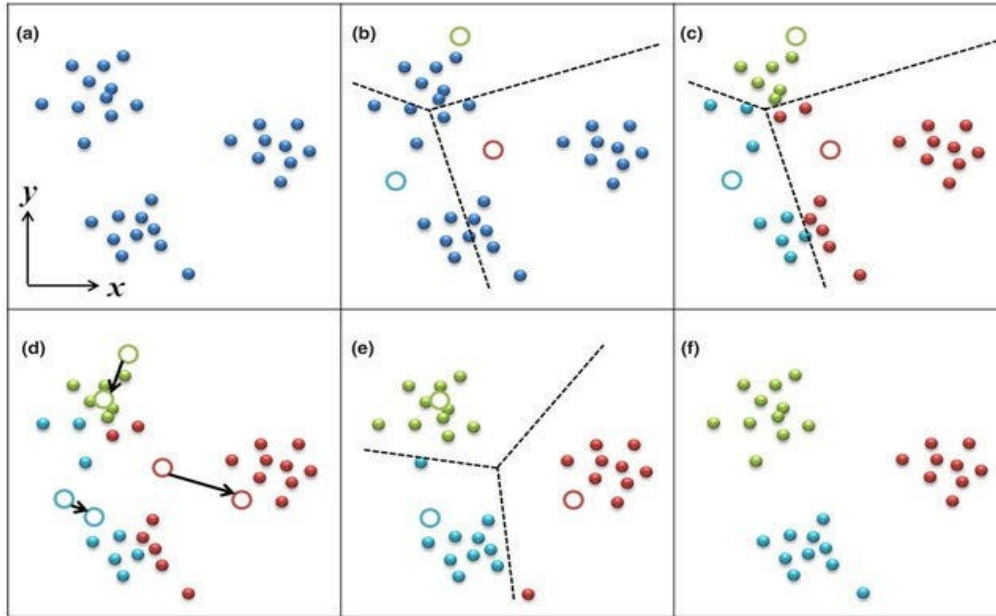
Agrupamento dos Dados (Clustering)

- Workflow:



Algoritmo K-Means

- Define grupos aproximando dados próximos aos 'centroides'
- Cada iteração ajusta os centroides e os dados pertencentes a cada grupo
- Ao fim, os centroides representam o ponto médio de cada cluster

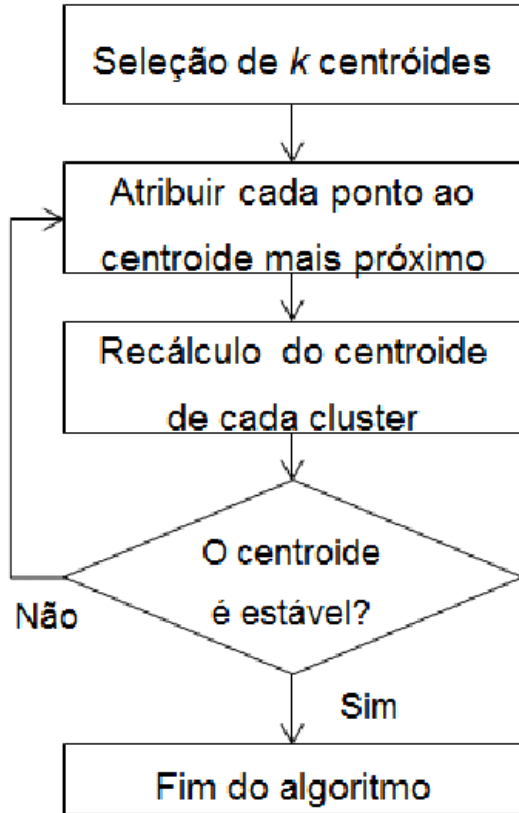


$$d = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}$$

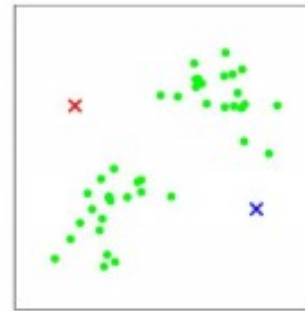
$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

Algoritmo K-Means

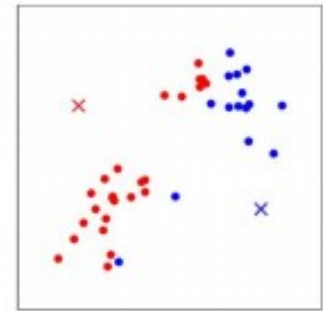
- Algoritmo



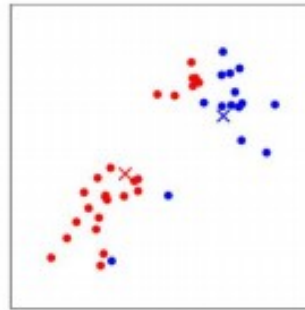
(a)



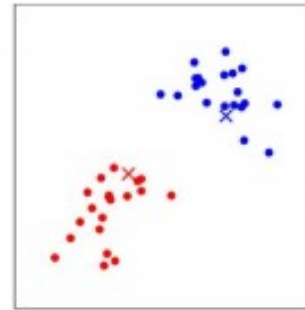
(b)



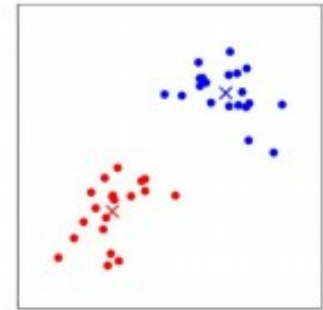
(c)



(d)



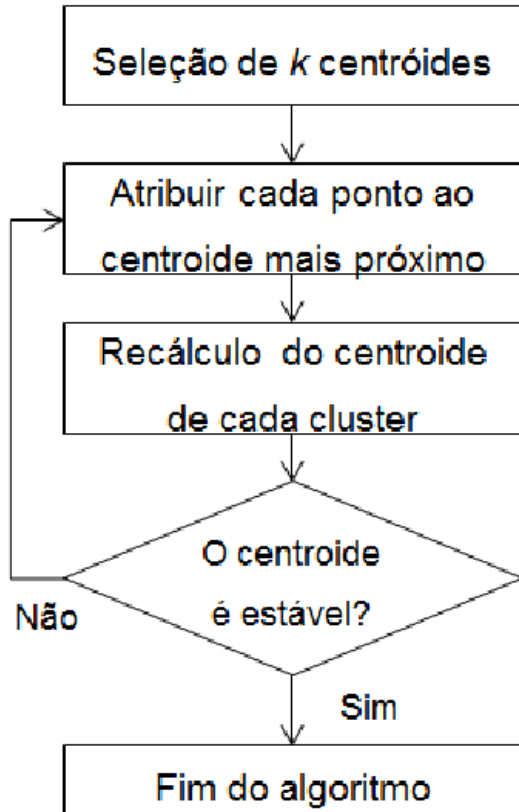
(e)



(f)

Algoritmo K-Means

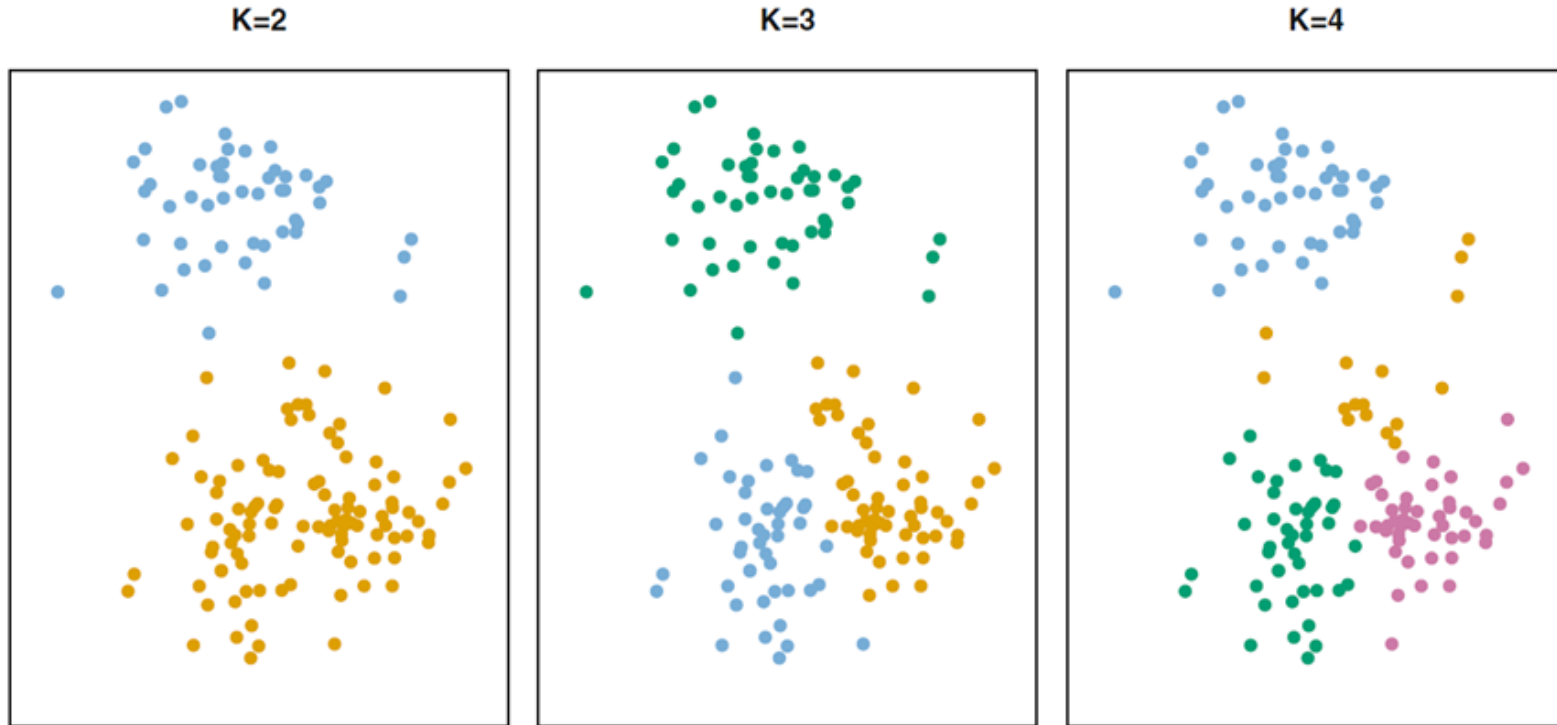
- Algoritmo



```
def fit(self, X, plot=False):  
    # Inicializa os centros dos clusters aleatoriamente  
    self.centers = X[np.random.choice(len(X), self.k, replace=False), :]  
  
    # Executa o algoritmo K-Means  
    for i in range(self.max_iters):  
        # Atribui cada ponto ao cluster mais próximo  
        distances = np.sqrt(np.sum((X - self.centers[:, np.newaxis]) ** 2, axis=2))  
        labels = np.argmin(distances, axis=0)  
  
        # Verifica se o algoritmo convergiu  
        old_centers = self.centers.copy()  
        for k in range(self.k):  
            self.centers[k] = np.mean(X[labels == k], axis=0)  
        if np.allclose(old_centers, self.centers, rtol=0, atol=self.tol):  
            print(f"Algoritmo convergiu após {i+1} iterações.")  
            break
```

Algoritmo K-Means

E como avaliar os clusteres ? Como determinar K?



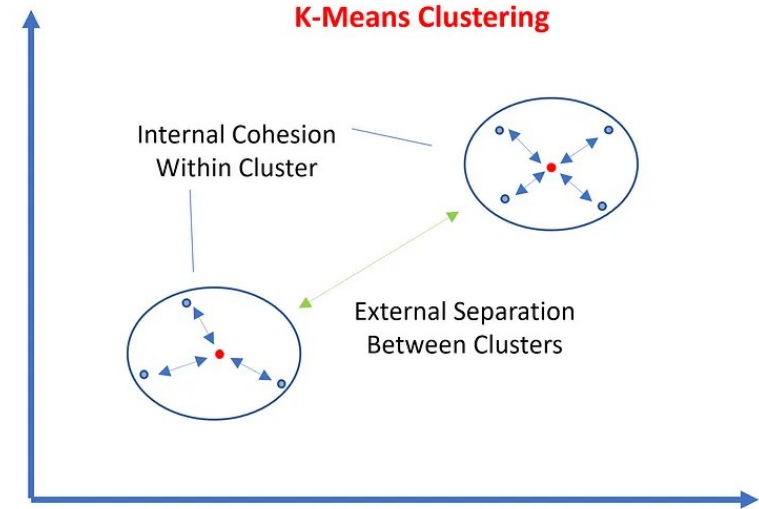
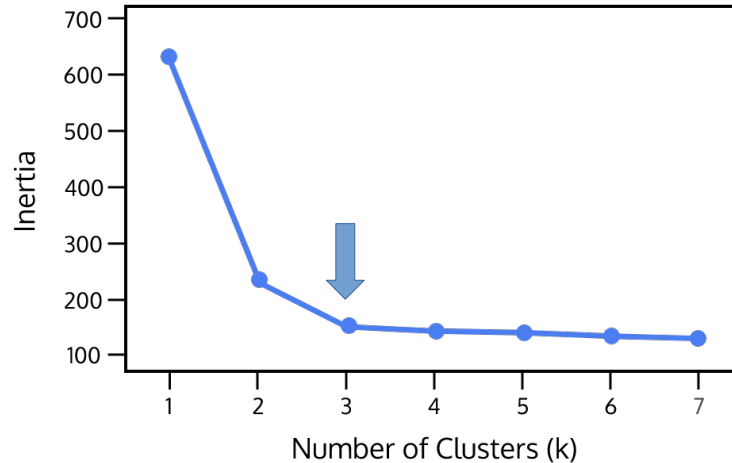
Algoritmo K-Means

- Inertia (Soma dos erros quadrados)

$$\sum_{i=1}^N (x_i - C_k)^2$$

- Elbow Method (Inertia global)

$$\sum_{j=1}^k \sum_{i=1}^N (x_i - C_k)^2$$

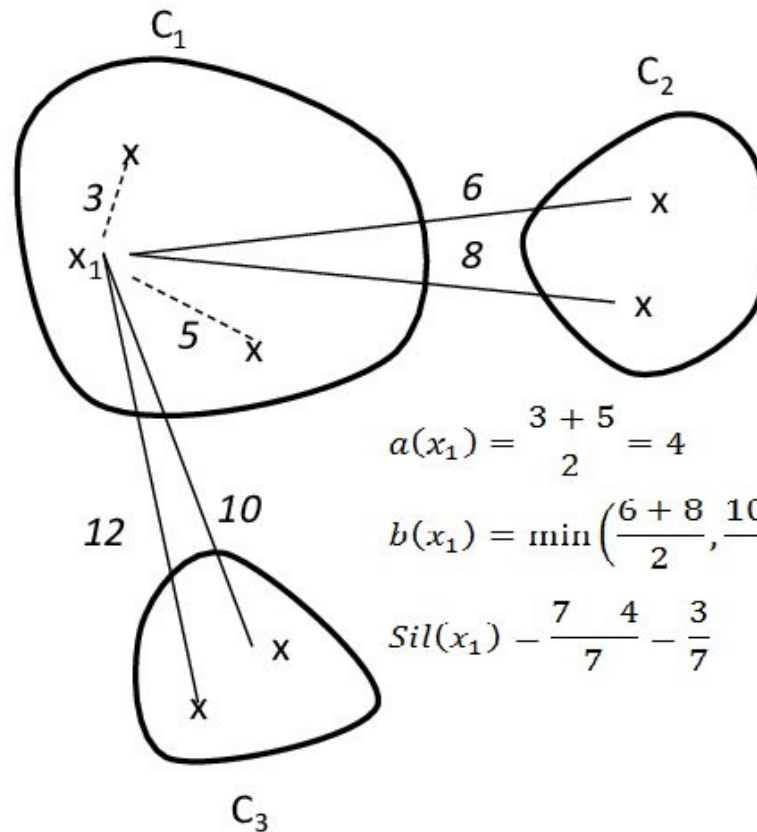


Algoritmo K-Means

- Silhueta (Silhouette)
- Medida por instâncias $[-1,1]$
- Coesão (a)
- Separação (b)

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

- $[-1,+1]$:
 - -1: Não coerente
 - +1: Coerente



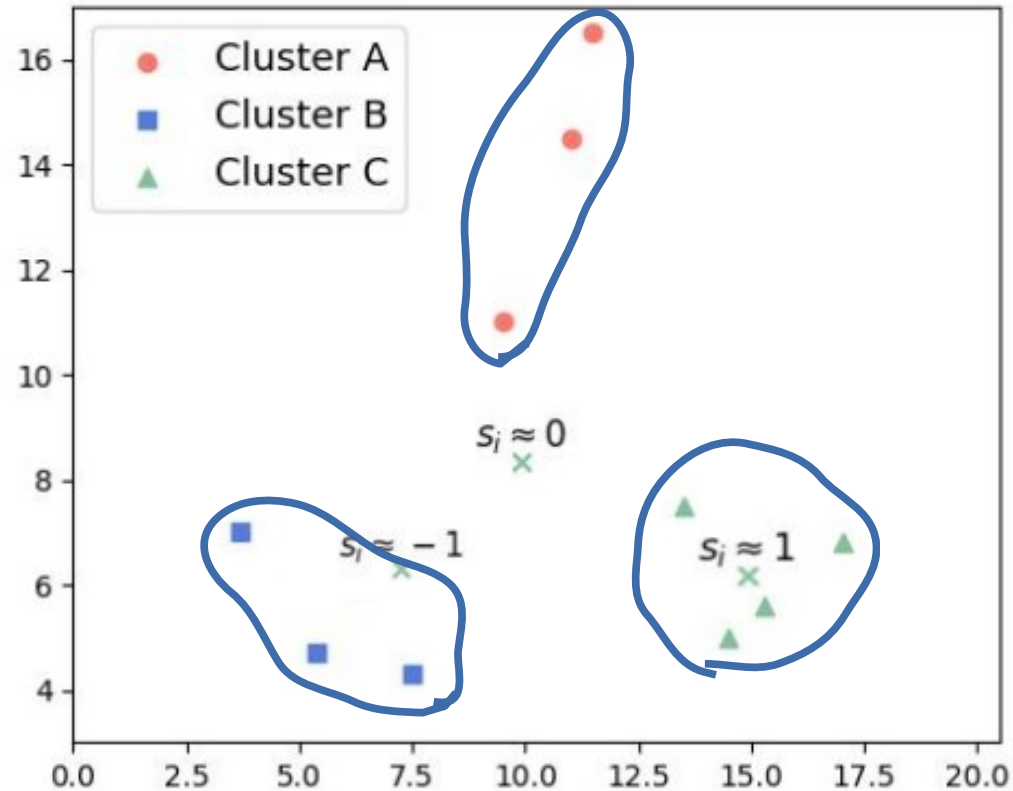
$$a(x_1) = \frac{3 + 5}{2} = 4$$

$$b(x_1) = \min\left(\frac{6 + 8}{2}, \frac{10 + 12}{2}\right) = 7$$

$$Sil(x_1) = \frac{7 - 4}{7} = \frac{3}{7}$$

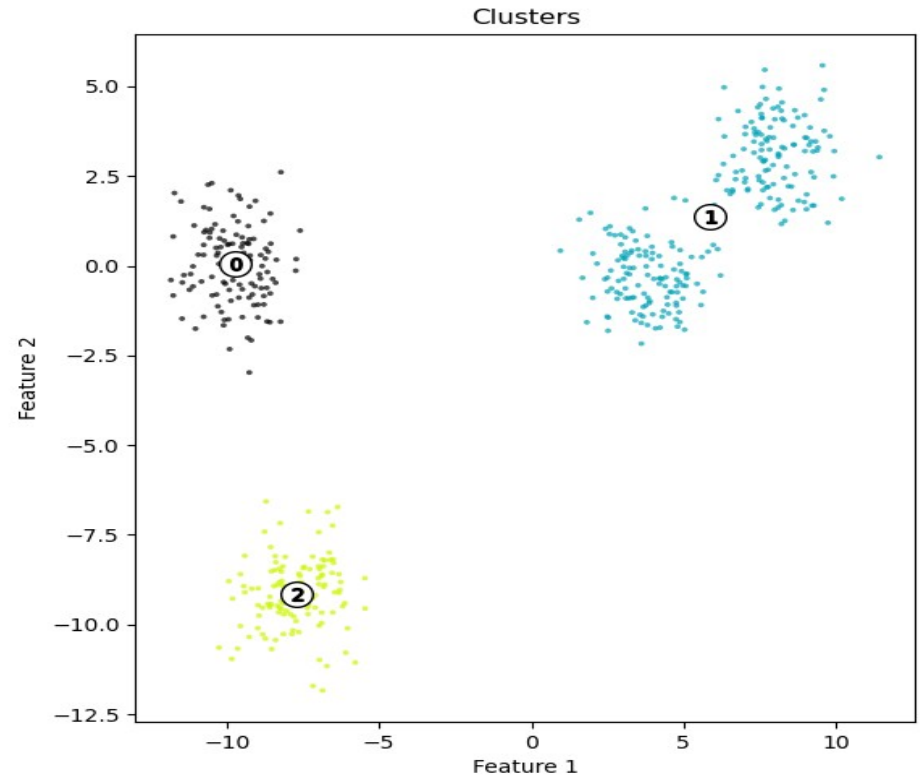
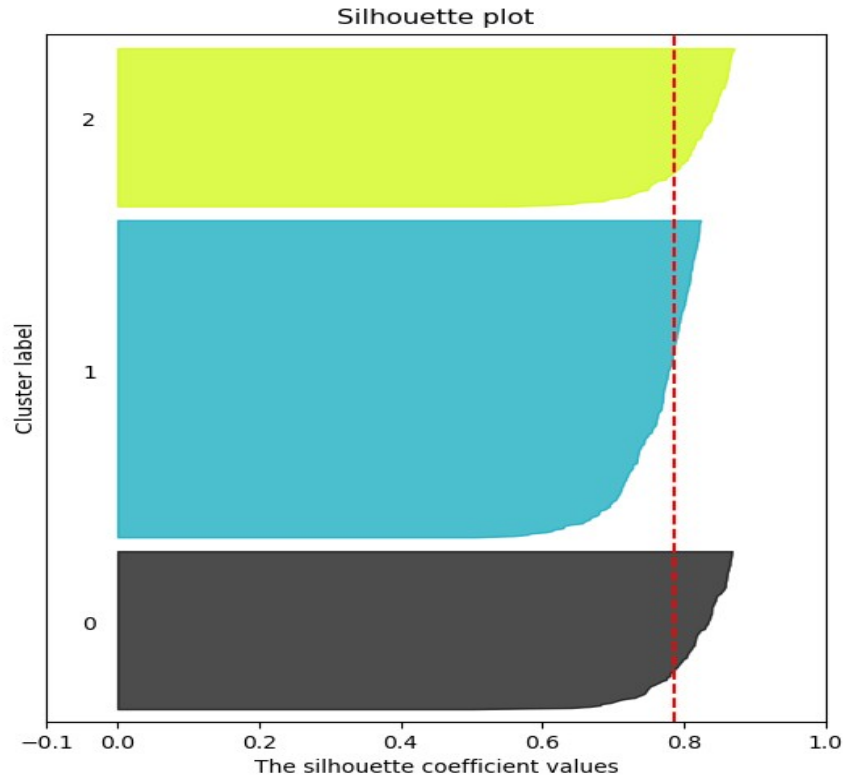
Algoritmo K-Means

- Silhueta



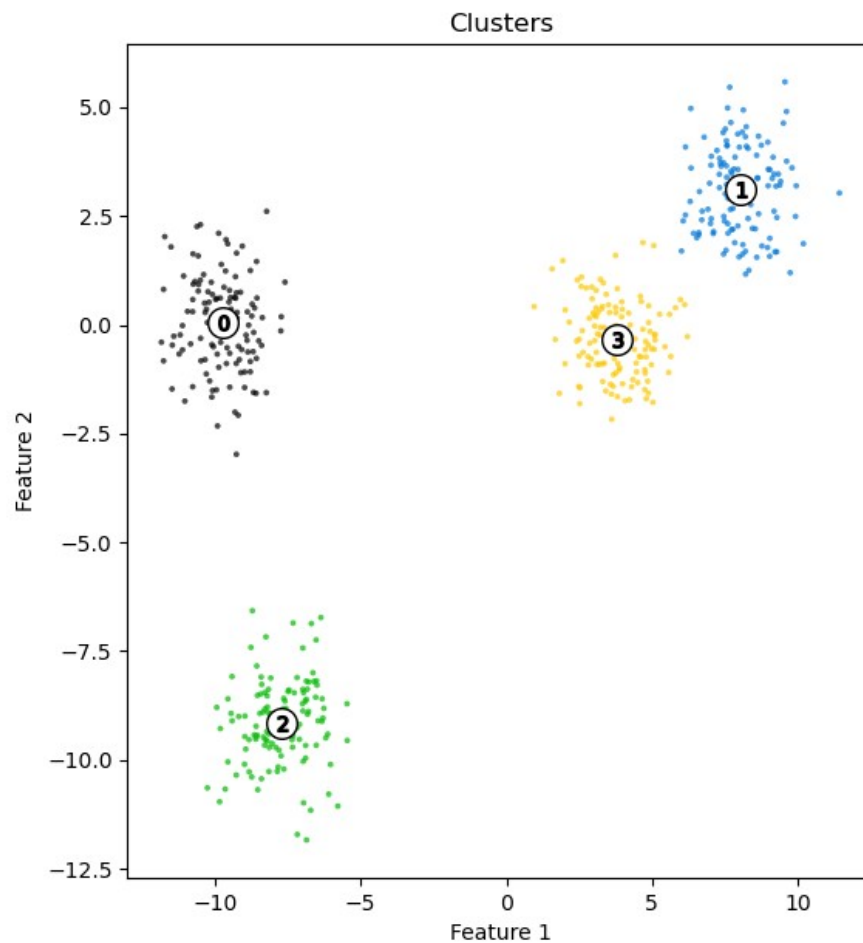
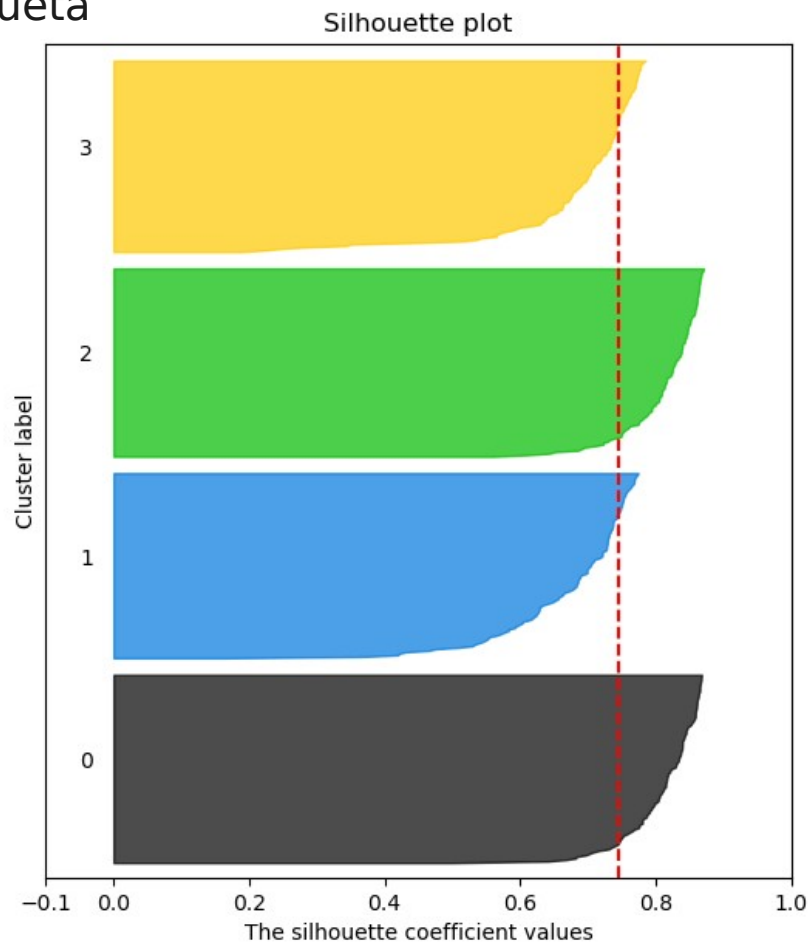
Algoritmo K-Means

- Silhueta



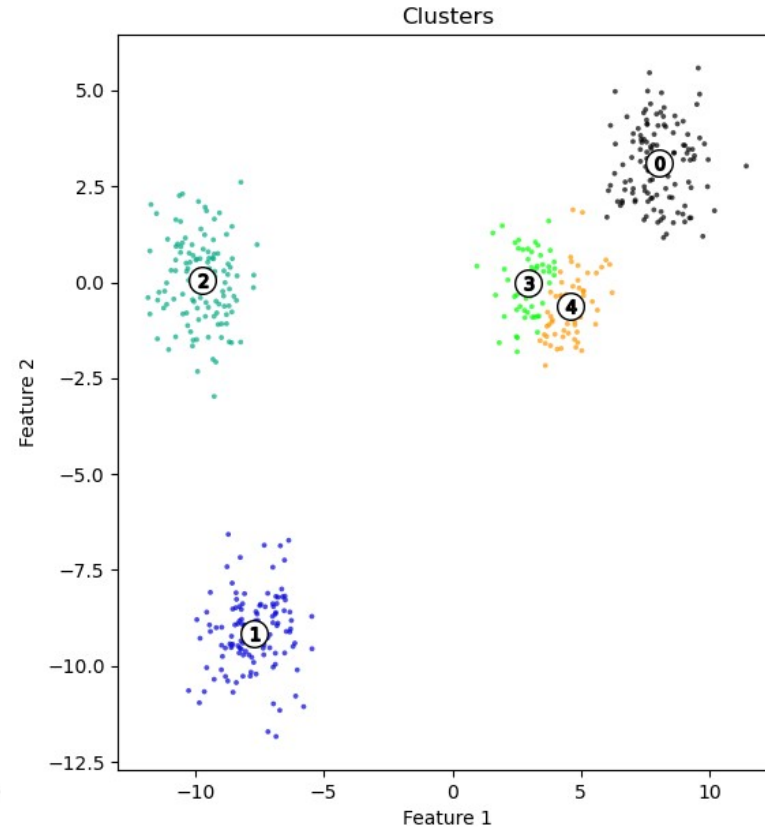
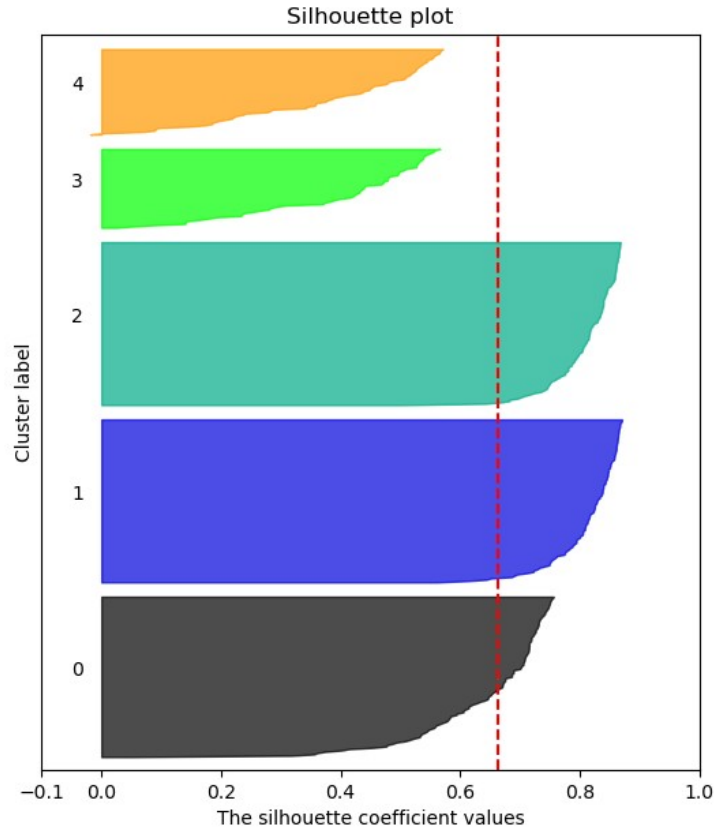
Algoritmo K-Means

- Silhueta



Algoritmo K-Means

- Silhueta



Normalização e Redução de Dados

- Normalização
 - Normaliza a escala dos dados
 - Importante para algoritmos que analisam a distribuição espacial das características
 - Realizada por atributos

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
297	59	1	0	164	176	1	0	90	0	1.0	1	2	1
243	57	1	0	152	274	0	1	88	1	1.2	1	1	3
269	56	1	0	130	283	1	0	103	1	1.6	0	0	3
215	43	0	0	132	341	1	0	136	1	3.0	1	0	3
83	52	1	3	152	298	1	1	178	0	1.2	1	0	3
152	64	1	3	170	227	0	0	155	0	0.6	1	0	3

Let's Code!

- LINK: ==> : Tutorial - Agrupamento

Normalização

MinMax (sklearn.preprocessing.MinMaxScaler)

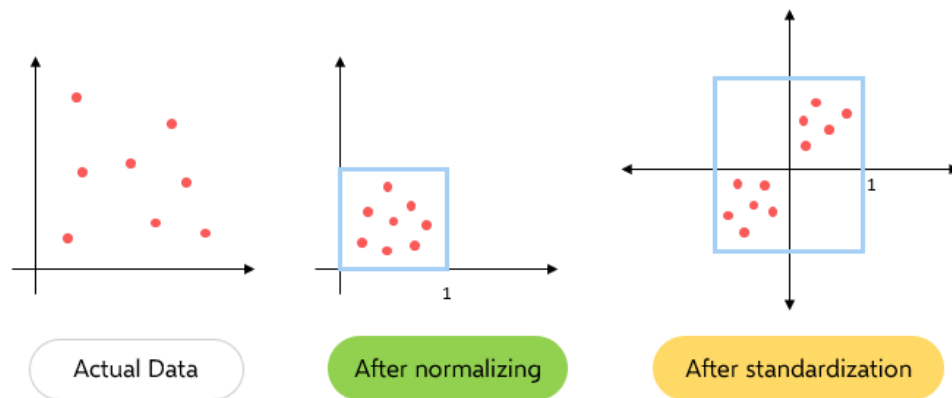
- 0 - 1
- Redes Neurais (i.e neurônio relu)
-

$$\frac{x_i - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})}$$

Std (sklearn.preprocessing.StandardScaler)

- -stdDev , 0 , +stdDev
- Métodos espaciais (K-nn, K-means, SVM)

$$\frac{x_i - \text{mean}(\mathbf{x})}{\text{stdev}(\mathbf{x})}$$



Normalização

MinMax (sklearn.preprocessing.MinMaxScaler)

$$\frac{x_i - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})}$$

```
Original: [ 171  216 -188  123  -61 -187   20 -161   33   -9]
Normalized: [[0.889 1.      0.      0.77  0.314 0.002 0.515 0.067 0.547 0.443]]
-----
Original: [ 37.038  33.629   6.81 -13.202 -13.249 -44.837  -7.96  20.623  11.133
  8.058]
Normalized: [[1.      0.958 0.631 0.386 0.386 0.      0.45  0.8   0.684 0.646]]
```

Std (sklearn.preprocessing.StandardScaler)

$$\frac{x_i - \text{mean}(\mathbf{x})}{\text{stdev}(\mathbf{x})}$$

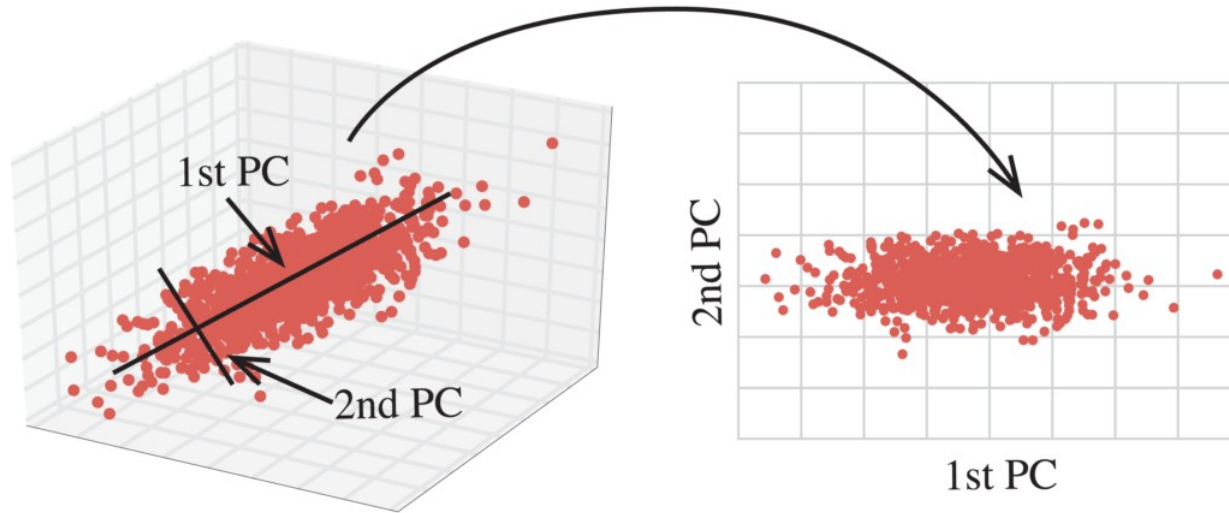
```
Original: [ 171  216 -188  123  -61 -187   20 -161   33   -9]
Normalized: [[ 1.264  1.588 -1.324  0.918 -0.409 -1.317  0.175 -1.13  0.269 -0.034]]
-----
Original: [ 37.038  33.629   6.81 -13.202 -13.249 -44.837  -7.96  20.623  11.133
  8.058]
Normalized: [[ 1.425  1.278  0.129 -0.729 -0.731 -2.085 -0.504  0.721  0.314  0.182]]
```

Redução de Dados

PCA calcula componentes principais que são combinações lineares das variáveis originais.

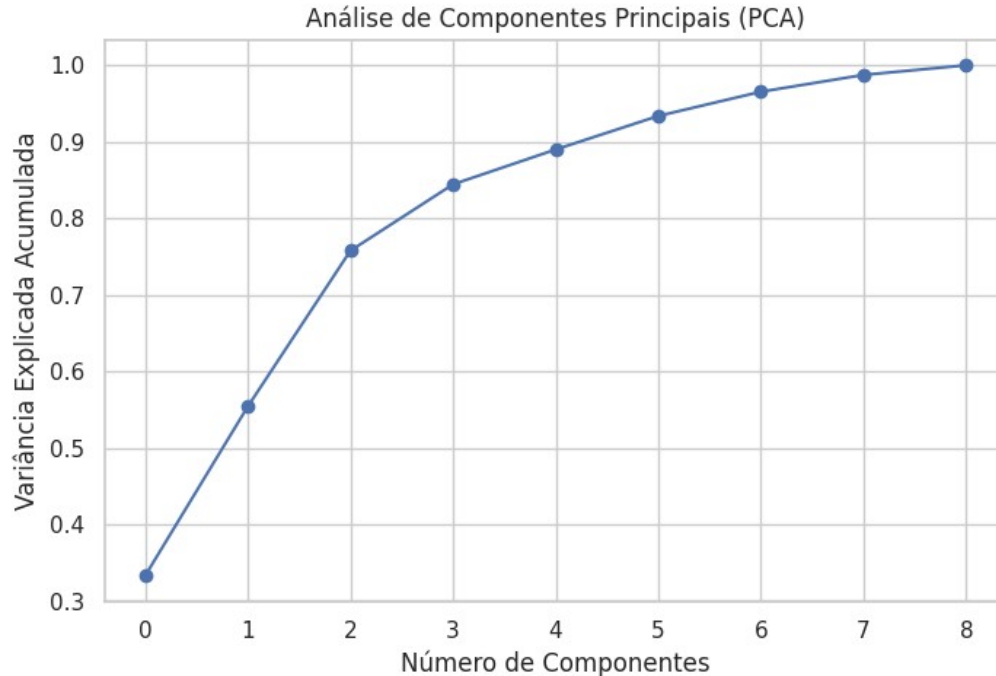
O método pode ser usado em momentos distintos:

- Visualização: Dado os diferentes atributos, como reduzir a sua complexidade para visualizar em um espaço 2D ou 3D



Redução de Dados

- Redução de Dados: Evita que dados redundantes, ruídos ou sem informação, sejam considerados no agrupamento



Let's Code!

LINK: ==> Tutorial - Normalização e Redução